Identification of natural mutations responsible for altered infection phenotypes of *Salmonella enterica* clinical isolates using cell line infection screens

Rafał Kolenda[1,2,#], Michał Burdukiewicz[3], Marcjanna Wimonć[1], Adrianna Aleksandrowicz[1], Aamir Ali[4], Istvan Szabo[5], Karsten Tedin[6], Josefin Bartholdson Scott[7], Derek Pickard[7], Peter Schierack[2,8]

Department of Biochemistry and Molecular Biology, Faculty of Veterinary Medicine, Wrocław University of Environmental and Life Sciences, Poland[1];

Institute of Biotechnology, Faculty Environment and Natural Sciences, BTU Cottbus-Senftenberg, Senftenberg, Germany[2];

Warsaw University of Technology, Warsaw, Poland[3];

National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan[4];

National *Salmonella* Reference Laboratory, German Federal Institute for Risk assessment (BfR), Berlin, Germany[5];

Free University Berlin, Berlin, Germany[6];

Cambridge Institute for Therapeutic Immunology & Infectious Disease, University of Cambridge Department of Medicine, Cambridge, United Kingdom[7];

Faculty of Health Sciences, Public Health Campus Brandenburg, Germany[8];

#Corresponding author:

Rafał Kolenda

Department of Biochemistry and Molecular Biology

The Faculty of Veterinary Medicine

Wrocław University of Environmental and Life Sciences

Norwida 31; 50-375 Wrocław

Telephone: +48713205209

e-mail: rafal.kolenda@upwr.edu.pl

Abstract

The initial steps of *Salmonella* pathogenesis involve adhesion and invasion of host epithelial cells. While well-studied for *S.* Typhimurium, the factors contributing to this process in other, host-adapted serovars remains unexplored. Here, we screened clinical isolates of serovars Gallinarum, Dublin, Choleraesuis, Typhimurium and Enteritidis for adhesion and invasion of intestinal epithelial cell lines of human, porcine, and chicken origins. 30 isolates with altered infectivity were used for genomic analyses and 14 genes and novel mutations associated with high or low infectivity were identified. The functions of candidate genes included virulence gene expression regulation, cell wall or membrane synthesis and components. The role of several of these genes in *Salmonella* adhesion and invasion to cells has not previously been investigated. The genes *dksA* (stringent response regulator) and *sanA* (vancomycin high-temperature exclusion protein) were selected for further analyses, and we confirmed their roles in host cell adhesion and invasion. Furthermore, transcriptomic analyses were performed for *S.* Enteritidis and *S.* Typhimurium, with two highly infective and two marginally infective isolates for each serovar. Expression profiles for the isolates with altered infection phenotypes revealed the importance of T3SS expression levels in the determination of isolate's infection phenotype. Taken together, these data indicate a new role in cell host infection for genes or gene variants previously not associated with adhesion and invasion to the epithelial cells.

Importance

Salmonella is a foodborne pathogen affecting over 200 million people and resulting in over 200,000 fatal cases per year. Adhesion to and invasion of *Salmonella* into intestinal epithelial cells is one of the first and key steps in the pathogenesis of salmonelosis. Still, around 35-40% of bacterial genes have no experimentally validated function and their contribution to the bacterial virulence, including adhesion and invasion, remains largely unknown. Therefore the

significance of this study is in the identification of new genes or gene allelic variants previously not associated with adhesion and invasion. It is well established that blocking adhesion and/or invasion would stop or hamper bacterial infection, therefore the new findings from this study could be used in future developments of anti-*Salmonella* therapy targeting genes involved in these key processes. Such treatment could be a valuable alternative as the numbers of antibiotic-resistant bacteria is growing very rapidly.

Keywords: *Salmonella*, adhesion, invasion, infection, clinical isolates, epithelial cells, *sanA*, *dksA*, genomics, transcriptomics

1. Introduction

Typhoidal and non-typhoidal salmonellosis is one of the major causes of foodborne diseases globally, affecting over 200 million people and resulting in over 200,000 fatal cases per year (1, 2). Animal salmonellosis is of considerable importance not only because of animal infection, but also as a reservoir for human infections (3). The causative agent in both humans and animals is *Salmonella enterica* subsp. *enterica*, which is comprised of more than 1500 serovars, of which only a few are of clinical significance (4). Based on clinical presentation and host-range, *Salmonella* serovars can be classified as either host-restricted, host-adapted and host-unrestricted (5). Most serovars from clinical samples are host-unrestricted and are able to infect multiple hosts, usually resulting in gastroenteritis (6). Host-adapted 'specialists' like *S.* Dublin and *S.* Choleraesuis mainly cause systemic disease in cattles and pigs, respectively, but can sporadically cause systemic infections in other hosts (e.g. human) (7, 8). Host-restricted serovars are limited to a single host where they cause systemic infections, for example *S.* Gallinarum (poultry) or *S.* Typhi (human and primates) (1, 9).

As a facultative intracellular, gastrointestinal pathogen, the initial steps of *Salmonella* pathogenesis involve adhesion to and invasion of host intestinal epithelial cells (10). *Salmonella* has evolved various strategies for adhesion and invasion, and different serovars have developed specific combinations of various adhesins or variants of the same adhesin for host cell attachment (11, 12). Adhesive structures of *Salmonella enterica* can be divided into three groups: fimbrial adhesins, non-fimbrial adhesins, and atypical adhesive structures (13). The role of the majority of these in the pathogenesis of various *Salmonella* serovars remains unknown. *Salmonella* invasion of host cells is mediated by the type three secretion system (T3SS) encoded on the *Salmonella* pathogenicity island 1 (SPI-1) (14), although additional factors such as the proteins Rck and PagN have also been found to play a role (15, 16). Despite an overwhelming amount of research and a seemingly clear picture of the adhesion and invasion process, largely based on work with the serovar *S.* Typhimurium, our understanding of these phenomena is still not complete. For example, it is now evident that adhesion and invasion can be affected by expression of different adhesin variants, mutations in genes encoding T3SS effectors, environmental factors, stress, metabolism, and various regulatory proteins (12, 17). Additionally, around 35-40% of bacterial genes have no experimentally validated function and their contribution to the bacterial virulence, including adhesion and invasion, remains unknown (18). Therefore the search for new genes involved in these processes is highly desirable.

To find new virulence factors contributing to adhesion or invasion, in this study we investigated the adhesion and invasion of various *Salmonella* serovars against model cell lines derived from three host species: human Caco-2, porcine IPEC-J2 and chicken CHIC-8E11. The study included host-restricted *S.* Gallinarum (from chicken), host-adapted *S.* Dublin (from cattle) and *S.* Choleraesuis (from various animals, largely swine) and host-unrestricted *S.* Typhimurium (from humans and pigs) and *S.* Enteritidis (from humans and chicken). For

101 quantification of adherent and invasive *Salmonella in vitro*, we used a previously established
102 VideoScan technology (19, 20) with a new VideoScan module specifically developed for
103 high-throughput quantification of *Salmonella* infected cells. Moreover, using whole genome
104 sequencing, we identified mutations affecting adhesion and invasion in novel genes, for which
105 contributions to adhesion and invasion were further validated. Additionally, we used RNAseq
106 to identify the differentially expressed genes affecting adhesion and invasion. This study
107 improves the understanding of mechanisms underlying adhesion and invasion of host
108 epithelial cells during *Salmonella* infection and shows contribution of small genome
109 alterations to the bacterial virulence.
110

111 **2. Materials & methods**

112 **2.1. Bacterial isolates and plasmids**

113 All bacterial isolates used in this study (Table 1) were tested for ampicillin (Amp) and
114 kanamycin (Kan) resistance with agar disc diffusion tests. For ampicillin-sensitive isolates,
115 plasmid pFPV25.1, containing an Amp resistance gene and the GFPmut3 fluorescent protein
116 expressed from the constitutive *rpsM* promoter, was used (21) . To exchange the Amp
117 resistance gene to a Kan resistance gene for those isolates that were ampicillin-resistant but
118 kanamycin-sensitive, Gibson assemble reaction was performed (NEB). The primers that were
119 used for the amplification of pFPV25.1 plasmid and Kan resistance cassette are listed in Table
120 2. PCR was performed with the use of Phusion High-Fidelity DNA Polymerase (Thermo
121 Scientific), according to the manufacturer's protocol. Gibson assemble was performed
122 according to the manufacturer's protocol. The new plasmid was named pFPV25.1Kan.

123 For adhesion and invasion screening, all 127 *Salmonella* isolates were GFP-labelled by
124 transformation with pFPV25.1 or pFPV25.1Kan. Electrocompetent bacteria were prepared as
125 described previously by (Sambrook and Russell, 2006) with some modifications. After

transformation with plasmid, single colonies were re-streaked on fresh Luria-Bertani (LB) agar plates (with appropriate antibiotics). Cultures were prepared in 1 ml of LB (with appropriate antibiotics) in 2-ml Eppendorf tubes and grown for 16 h, and glycerol-based cryostocks were prepared from these cultures.

2.2. Cell culture

Three intestinal epithelial cell lines were used in this study: porcine IPEC-J2 (obtained from Peter Schierack, (22)), human Caco-2 (DMSZ, Germany) and chicken CHIC-8E11 (MicroMol GmbH, obtained from Karsten Tedin, FU Berlin). Cells were grown in D-MEM/Ham's F-12 (Millipore) supplemented with 1 mM L-glutamine and penicillin/streptomycin (Millipore) and either 5% (IPEC-J2 and CHIC-8E11) or 10% (Caco-2) bovine serum (Millipore), and were incubated at 37˚C with 5% CO_2 and passaged using standard protocols. For infection assays IPEC-J2, Caco-2 and CHIC-8E11 were seeded in Nunclon Delta 96-well plates (Nunc) at a density of $0.7x10^4$, $1x10^4$ and $2x10^4$, respectively and used after 5 days. For gentamycin protection assay, Caco-2 cells were seeded in 24-well plates (Nest) at a density $5.8x10^4$ and used after 5 days.

2.3. VideoScan module for cell line infection assays

For automated bacterial quantification, we further developed our VideoScan technology (23) by creating a new module. Briefly, the VideoScan hardware consists of an inverse epifluorescence Olympus microscope with multiple fluorescence channels, a digital camera and a motorized scanning stage. The VideoScan module works with a 20x magnification objective and first focuses on DAPI-stained cell nuclei in a well, which are then marked as position 0 nm. Afterwards the software takes a Z-stack of images (for IPEC-J2, CHIC-8E11, 6 images; Caco-2, 7 images) of bacteria (GFP stained) with a starting position of -2000 nm (for IPEC-J2, CHIC-8E11) or 0 nm (for Caco-2). Next, one composite image is made from the Z-

150 stack images taken and bacteria on the image are detected and counted. Twenty (20)
151 composite images per well were analysed during read-out.
152
153 **2.4. Determination of the linear range of the VideoScan assay**
154 To test the reliability of our automated bacteria quantification assay, *S.* Typhimurium SL1344,
155 transformed with a pFPV25.1Kan plasmid, was grown to late exponential phase (OD_{600}=0.95-
156 1.0, 2.85-3x10^8 of bacteria/ml) at 37˚C and 180 rpm and washed with Phosphate-Buffered
157 Saline (PBS). After this, 50 µl of bacteria were applied in a dilution series (2x10^4-3x10^7/well;
158 final concentration of bacteria- 4x10^5-6x10^8/ml) on plates with a monolayer of IPEC-J2,
159 Caco-2 and CHIC-8E11 cells. Bacteria were incubated for 1 h and the plates were washed and
160 the underlying cell line monolayer with bound bacteria was fixed with 4% paraformaldehyde
161 (PFA, Roth). Cell nuclei were stained with DAPI. Next, adhered bacteria in the plates were
162 counted using the VideoScan module. Three independent experiments with three repetitions
163 for each dilution were prepared and measured.
164
165 **2.5. Cell line infection assays**
166 To screen the 127 *Salmonella* isolates for infectivity, cell lines were seeded in 96-well
167 plates and assayed on day 5, 6 and 7 after seeding. 127 *Salmonella* isolates with GFP
168 expression were used for infection assays. *S.* Typhimurium SL1344 was used as a standard
169 strain in all infection assays. Bacteria were grown in 1 ml of LB (with 50 µg/ml of Kan or 100
170 µg/ml of Amp) in 2-ml Eppendorf tubes at 37°C and 180 rpm. Before an assay, bacteria were
171 diluted to a concentration of 4x10^7 bacteria/ml and 50 µl of bacteria were applied per well to a
172 monolayer of IPEC-J2, Caco-2 and CHIC-8E11 cells. Bacteria were incubated for 1 h and 4 h
173 in a cell culture incubator and the plates were washed three times with PBS. Cell lines with
174 bound bacteria were fixed with 4% PFA in PBS, and the plates were washed with PBS and

cell line nuclei were stained with DAPI. Afterwards, adhered bacteria were counted with the VideoScan. Three independent experiments with three repetitions for each *Salmonella* isolate was performed.

2.6. Next-generation genome sequencing and comparative genome analysis

Genomic DNA was isolated with the Wizard® Genomic DNA Purification Kit (Promega) according to the manufacturer's instruction. DNA was sequenced with the use of HiSeq X platform at Sanger Institute Sequencing Facility. Sequencing quality was assessed on the basis of average base quality, GC content and adapter contamination (24). Genomes were assembled using the Shovill pipeline and assembly_improvement pipeline (25). Genome assemblies were annotated with Prokka (26). Pangenomes were determined by using Roary (27). Core genome phylogeny was performed with FastTree 2.1 software using generalized time-reversible model of nucleotide evolution. The FastTree 2.1 produces results similar to Maximum Likelihood analysis and the accuracy and reproducibility of this method has been estimated (28). Phylogenetic trees were annotated with iTOL software (29). Comparison of gene presence between isolates with different infection phenotypes was performed with Scoary (30). Multilocus sequence types were determined with the MLST 2.1 online tool and the presence of pSV plasmid was checked with the PlasmidFinder (31, 32).

For single nucleotide polymorphism search, the reads obtained for each isolate were mapped against a reference genome using bwa (33) (Supplementary Table 1). Mapped reads were individually analysed with Artemis (34). All isolates belonging to the same serovar were compared with a reference genome from the same serovar. The mutation found in each genome was considered as potentially associated with altered infection phenotypes if it was deletion or single nucleotide substitution which resulted in premature stop codons and/or was placed in the gene with confirmed contribution to adhesion or invasion.

200

201 **2.7. Bacterial mutant construction**

202 For *S.* Enteritidis clinical isolates no. 5706 and 6203, scarless and markerless negative
203 selection-based system was utilized to cure the identified mutations of interest as in the
204 publication of Khetrapal et al. (2015) (35). Primers used for mutant generation are listed in
205 Table 2. To determine whether the isolates showed altered morphologies in comparison to the
206 parental isolates, cultures (grown for 16 h) were stained with acridine orange and investigated
207 with the use of a fluorescence microscope.

208 **2.8. Growth curves**

209 Overnight (O/N) cultures (grown for 16 h) or exponential growth cultures were diluted to
210 $5x10^6$ bacteria/ml in LB, antibiotic-free cell culture medium or M9 minimal medium with 0.2
211 % glucose (for strain 5706) and 200 µl of bacterial suspension was pipetted to each well of
212 96-well polystyrene plate. Bacterial growth was measured with a Tecan plate reader for 16 h
213 with 30 seconds of vigorous shaking and OD_{600} measurement every 15 minutes. Three
214 technical and biological repetitions were performed for each strain. Negative controls were
215 included for each medium used in each assay. To asses differences in growth of 5706 isogenic
216 strains, the package Growthcurver was used and growth rate (r), doubling time (t_gen) and
217 empirical area under curve (AUC) were used for comaprisons between strains (36).

218 Growth with vancomycin for strain 6203 was performed with the use of O/N cultures
219 (grown for 16 h). Bacteria were diluted in Müller-Hinton Broth (MHB, Roth) to concentration
220 of $1x10^6$ bacteria/ml. Two-fold serial dilutions of vancomycin in 50 µl of MHB were prepared
221 in 96-well polypropylene plates with starting concentration of vancomycin set to 1 mg/ml.
222 Afterwards, 50 µl of bacterial suspension was pipetted into each well. Bacterial growth was
223 measured with the use of a Tecan plate reader for 16 h with 30 seconds of vigorous shaking

and OD_{600} measurement every 15 minutes. For each strain 3 technical and biological repetitions were performed. No-antibiotic and negative controls were included in each assay.

2.9. RNA isolation and real-time quantitative PCR (qPCR)

Overnight cultures of strains 5706 and 6203 and their isogenic mutants were grown to OD_{600}= 0.45-0.55. Next, 1 ml of bacteria was harvested and RNA was isolated by the RNAsnap method (37). For each sample, 1 µg of RNA was digested with DNase I (Thermo Scientific) according to the manufacturer's instructions and reverse transcription was performed with the use of iScript cDNA Synthesis Kit (BioRad). qPCR was performed with the use of Universal SYBR Green qPCR Supermix (BioRad) and CFX96 Thermocycler (BioRad). Primers used for qPCR are listed in Table 2. The qPCR data analysis was performed with CFX Manager 3.1 (BioRad) by normalizing the expression of *dksA* and *sanA* genes to the reference gene *tufA* with ΔΔCq method (38).

2.10. Gentamycin protection assays

Adhesion and invasion assays were performed on strains generated in section 2.8. We chose isolates with an altered infection phenotype. Here, we compared isolates with high infectivity (=one phenotype) with isolates with low infectivity (=another phenotype) within one serovar. As such altered infection phenotypes were similarly identified for Caco-2, IPEC-J2 and CHIC-8E11 cells, and further work was done only with the Caco-2 cell line.

All bacteria were previously determined to be susceptible to gentamycin at a concentration of 50 µg/ml. First, O/N cultures of bacteria (grown for 16 h) in 1 ml of LB in 2-ml Eppendorf tubes were diluted to OD_{600}=0.05 in 5 ml of LB in 15-ml falcon tubes and grown to OD_{600}=2.0. Next, bacteria were washed once in PBS and resuspended in antibiotic-free cell culture medium. Bacteria ($1.2x10^7$) in 1 ml of medium was added to Caco-2 cells per well (proportional to the concentration of bacteria in infection assays in a 96-well plate format).

After 1 h incubation of bacteria with cells, bacterial suspension was discarded, cells were washed three times with 1 ml of PBS and new medium with or without gentamycin was added. After 1 h of incubation, supernatants were discarded and cells were washed three times with PBS. Cell monolayers were lysed with 1% of Triton-X 100 (Sigma) in PBS, dilution series in PBS were made and bacteria were spread on LB agar plates. Next morning, colony-forming units were counted. For each strain at least 3 technical and biological repetitions were performed.

2.11. RNA sequencing and differential transcriptome analysis

Four *S.* Typhimurium and four *S.* Enteritidis isolates with lowest and highest infectivity were selected for RNA sequencing. Overnight cultures were diluted 100x in 10 ml of LB in 50-ml Falcon tubes and grown to OD_{600}=0.4. Next, 5 ml of bacterial culture was added to 10 ml of RNAprotect Bacteria (Qiagen), vortexed 10 sec. and incubated at room temperature for 5 min. Next, all tubes were centrifuged 4,400xg for 15 min. at room temperature, supernatants were discarded and bacterial pellets were frozen at -80 ˚C (the pellet was stored and used for no longer than 1 week).

RNA isolation was done with RNeasy Mini Spin Kit (Qiagen). Next, RNA was digested with RNase-free DNase (Qiagen). RNA quantity and purity was checked with Nanodrop and RNA quality was determined with Agilent 2100 Bioanalyzer.

RNA was sequenced with the use of HiSeq X platform at Sanger Institute Sequencing Facility. Library for sequencing was prepared with ‘NEB Ultra II RNA Custom Kit’ (NEB). Sequencing quality was assessed on the basis of average base quality, GC content and adapter contamination (24). Reads were mapped to the reference genomes of *S.* Typhimurium SL1344 and *S.* Enteritidis P125109, respectively, and counted with the use of Rsubread package (39). Differential expression analysis was performed with use of edgeR (40), and a gene was considered as differentially express (DE) when FDR value was equal or lower than 0.05.

2.12. Data processing and statistical analysis

Statistical analysis was performed using R software (41). All figures were prepared with the ggplot2 package implemented in the R software (42).

For screening of the 127 *Salmonella* isolates results from three experiments (in total nine measurements) were baselined using results from empty-well controls. First, to eliminate the inter-replicate variance, median infection value (bacteria/mm^2) for each of three replicates from one measurement was computed. Next, the final infection value (bacteria/mm^2) was computed as a median of medians coming from three experiments. To facilitate comparison between measurements from different plates, final median values were standardized using the median value of SL1344 strain for a given plate. The procedure was repeated for three cell lines and both incubation periods. Standardized data were later compared using Mood test with Benjamini-Hochberg correction for multiple comparisons. Statistical analysis for growth curves and gentamycin protection assays was performed using Wilcoxon test implemented in R software.

2.13. Data availability

Genome sequence and RNAseq data for *Salmonella enterica* have been deposited at the NCBI under the BioProject ID PRJNA626643 and PRJNA667254, respectively.

3. Results

3.1. *Salmonella* transformation

Bacterial isolates were transformed with a GFP-expressing plasmid for automated quantification by our VideoScan method. To do this, susceptibility to ampicillin and kanamycin was first assessed for 128 *Salmonella* isolates. Most isolates were sensitive to both antibiotics, one isolate was resitant to kanamycin, and 37 were resistant to ampicillin. A single isolate was resistant to both ampicillin and kanamycin and was therefore excluded from

300 further work. Due to the high number of ampicillin-resistant isolates, the antibiotic resistance
301 cassette in plasmid pFPV25.1 was exchanged from ampicillin to kanamycin to create plasmid
302 pFPV25.1Kan. The remaining 127 isolates were successfully transformed with this new
303 plasmid.
304
305 **3.2. Quantification of *Salmonella* with VideoScan**
306 Before proceeding to automated quantification of infectivity of the 127 GFP-expressing
307 isolates, the reliability of the fluorescence microscope-based VideoScan method was assessed.
308 Serial dilution experiments were performed with *S.* Typhimurium SL1344 to determine the
309 dynamic range of the VideoScan. The quantification of *Salmonella* was possible within a
310 wide range of bacterial dilutions for all three cell lines tested (Fig. 1). The linear range of the
311 assay was $2x10^4$-$2.5x10^7$ (R^2=0.96), $2x10^4$-$3x10^7$ (R^2=0.96) and $2x10^4$-$3x10^7$ (R^2=0.92) for
312 IPEC-J2, Caco-2 and CHIC-8E11, respectively. Thus, quantification of GFP-labelled
313 *Salmonella* was reliable within the tested serial dilution range. A bacterial concentration of
314 $2x10^6$ bacteria/well ($4x10^7$/ml) was chosen for further assays.
315
316 **3.3. Cell line infection screening revealed no link between cell type and source of**
317 **host-specialist *Salmonella* isolates**
318 Porcine (IPEC-J2), human (Caco-2) and chicken (CHIC-8E11) intestinal epithelial cells were
319 infected with the 127 GFP-expressing *Salmonella* isolates for 1 and 4 h. Bacteria which
320 infected cells were counted using the VideoScan and data were fitted to an appropriate
321 statistical model (Fig. 2).
322 Surprisingly, *Salmonella* host-specialists did not appear to be more infective in cell lines
323 originating from the host species they were isolated from. *S.* Choleraesuis (pig specialist)
324 showed 12.5 times lower infection rates in porcine IPEC-J2 cells than Caco-2 ($p<0.01$),

325 whereas *S.* Gallinarum (chicken specialists) showed the highest infection in these cells, and
326 this result was 20 times higher than infection in CHIC-8E11 cells ($p<0.001$). Also, *S.*
327 Typhimurium isolated from humans did not show greater infection of human Caco-2 cells.
328 Similar results were found using *S.* Enteritidis from humans or chicken and *S.* Typhimurium
329 from pigs. Among all *Salmonella* groups, *S.* Gallinarum was 50-100 times less infective than
330 other groups on CHIC-8E11 ($p<0.001$) and *S.* Choleraesuis was 1.2-28 times more infective
331 than other groups on Caco-2 (statistically significant only for comparison with *S.* Gallinarum,
332 $p< 0.001$). In general, CHIC-8E11 cells were most susceptible to *Salmonella* infection
333 ($p<0.01$), while IPEC-J2 were most resistant ($p<0.01$) (Supplementary Table 2).
334 When the increase of infecting bacteria was calculated between the two incubation times
335 ($\Delta Inf = Inf_{4h} - Inf_{1h}$), *S.* Choleraesuis had the highest median increase (1.2-33-fold) on Caco-2
336 cells when compared with other groups (statistically significant only for comparison with *S.*
337 Gallinarum, $p< 0.001$) (Supplementary Fig. 1, Supplementary Table 3). The highest ΔInf on
338 CHIC-8E11 cells was observed for *S.* Dublin (1.2-78-fold higher, statistically significant for
339 comparison with *S.* Gallinarum and *S.* Enteritidis human, $p< 0.01$) and on IPEC-J2 for *S.*
340 Typhimurium isolated from humans (1.3-8.3-fold higher, statistically significant in
341 comparison with *S.* Choleraesuis, *S.* Dublin, *S.* Enteritidis from chicken and *S.* Enteritidis
342 from human, $p< 0.01$). Nearly all *Salmonella* groups had the lowest ΔInf in IPEC-J2 with the
343 exception for *S.* Gallinarum, which had the highest ΔInf in this cell line ($p<0.01$).
344
345 **3.4. Analysis of *Salmonella* genomes revealed the presence of natural mutations**
346 **responsible for altered infection phenotypes**
347
348 To identify bacterial factors responsible for varying infectivity, the genomes of 30 isolates
349 with the highest and lowest infectivity were sequenced and compared. The pangenome

350 analysis revealed no differences in gene content (i.e. gene presence or absence) associated
351 with infection phenotype between highly infective isolates in comparison to marginally
352 infective isolates. To determine the similarity among isolates with altered infection
353 phenotypes, core genome-based and additional binary genome-based phylogenetic trees were
354 constructed (Fig. 3, Supplementary Fig. 2). Moreover, MLST sequence types (ST) and the
355 presence of pSV plasmids which are common in almost all serovars were determined. Highly
356 infective isolates did not form separate subclades/groups and did not belong to different STs.
357 Presence of pSV plasmids did not correlate with infectivity levels.

358 All the results at this stage of work lead to the hypothesis that altered infection phenotypes
359 were driven not by gene content, but rather sequence variation in *Salmonella* genomes.
360 Therefore sequencing reads were mapped to reference genomes and aligned reads were
361 compared to identify any gene mutations in comparison to the reference genome. Fourteen
362 candidate genes were found to be potentially associated with altered infection phenotypes and
363 are listed in Table 3 alongside with function, description of mutation and whether similar
364 mutations could be found in genomes in the GenBank database. The functions of these 14
365 candidate genes were: 1) virulence gene expression regulation (e.g. *dksA*), 2) synthesis of cell
366 wall or membrane elements (e.g. *rfaL*), 3) and cell membrane components (e.g. *ompD*). The
367 role of several of these genes in *Salmonella* adhesion and invasion of cells has not been
368 investigated so far (e.g. *sanA* and *yidR*). Natural mutations found in the strains with altered
369 infection phenotypes included deletions, and single nucleotide substitution which resulted in
370 premature stop codons or amino acid alterations. Alleles with these natural mutations were
371 found for six genes in GenBank Reference Genomes database with BLASTN search.
372

373 **Functional assays confirm the role of *dksA* and *sanA* in host cell infection**

374 To confirm the contribution of mutations to altered infection phenotypes, two genes were
375 selected and genetically manipulated for further elucidation of their functions: *dksA* and *sanA*
376 (the function of each gene is summarized in Table 3). To study the effect of mutations found
377 in this work, *dksA* and *sanA* in *S.* Enteritidis clinical isolates (no. 5706 and 6203) were
378 replaced with the wild-type sequences from *S.* Enteritidis PT4 strain no. P125109. No obvious
379 phenotypic changes in comparison to parental strains were observed in the complemented
380 strains, indicating that the genetic manipulations did not affect bacterial morphology. Isogenic
381 strains of isolates 5706 and 6203 were tested for *dksA* and *sanA* expression, respectively.
382 Relative expression of wildtype *dksA* was 40% lower than the *dksA* mutated variant ($p = 0.001$). For 6203 isogenic strains, there was no difference in expression between wildtype and
383 0.001). For 6203 isogenic strains, there was no difference in expression between wildtype and
384 mutated *sanA* variants.
385 Gene *dksA* encodes a stringent response regulator required for growth in minimal medium,
386 therefore the isogenic strains of isolate 5706 were tested for growth in M9 minimal medium
387 supplemented with glucose. The strain expressing wildtype DksA had 1.7-1.9 faster growth
388 rate and doubling time in exponential phase of growth as compared to the strains expressing a
389 DksA mutated variant ($p< 10^5$, Figure 4 A, Supplementary Table 4). Subsequently, the ability
390 to adhere to and invade Caco-2 cells was evaluated and the wildtype strain had 80 % higher
391 infection and 30 % higher invasion levels compared to the strains with a *dksA* mutated variant
392 ($p = 0.05$) (Figure 4 B).
393 The *sanA* gene deletion in *E. coli* has been linked to increased susceptibility to
394 vancomycin. The function of this gene is often studied *in vitro* by vancomycin susceptibility
395 testing. Therefore, isogenic strains of isolate 6203 with wildtype and mutated variant of *sanA*
396 were tested for growth in vancomycin. Strains carrying the mutated variant were more
397 resistant to vancomycin than strain carrying the wildtype *sanA*. Strain with wildtype *sanA*
398 gene showed growth in up to 31.25 µg/ml vancomycin, whereas strains with a mutated *sanA*

399 survived concentrations up to 62.5 µg/ml. Next, strains carrying isogenic variants of *sanA*
400 were tested in adhesion and invasion assays. Strains carrying a mutated variant of *sanA*
401 infected cells 10 % higher and invaded 20 % higher than the strain carrying wildtype *sanA*
402 variant, but the statistical analysis revealed that these differences are not significant (Figure
403 4C).
404
405 **3.5. Different expression levels of T3SS are responsible for altered infection**
406 **phenotypes in *S.* Typhimurium and *S.* Enteritidis isolates**
407
408 Based on the results from the RNAseq and transcriptome studies, the number of differentially
409 expressed (DE) genes between isolates varied from 15 to 67 in the case of *S.* Enteritidis and
410 from 48 to 324 for *S.* Typhimurium isolates (Supplementary Table 5). All DE genes were
411 grouped by biological processes they represent and this analysis revealed the presence of two
412 gene groups- "Bacterial secretion system" and "Bacterial invasion of epithelial cells"-
413 associated with invasion of epithelial cells (Supplementary Figure 3A and 3B). These genes
414 belong to the T3SS injectosome machinery or are secreted into epithelial cells and mediate
415 bacterial internalization. Two *S.* Typhimurium isolates representing the high-infecting
416 phenotype (isolate no. 5727, 6191) had from 2- to 280-fold increased expression of these
417 genes in comparison to two isolates with a low-infecting phenotype (isolate no. 5735, 6185)
418 (Figure 5). This was not the case for *S.* Enteritidis, as one *S.* Enteritidis isolate (isolate no.
419 6203) with low-infecting phenotype had expression profiles similar to two *S.* Enteritidis
420 isolates with high-infecting phenotypes, indicating a contribution of other factors to the
421 adhesion and invasion process.
422 **4. Discussion**

423 This study was aimed at identifying new genes with possible contributions to adhesion
424 and/or invasion by *Salmonella* with host epithelial cells. To develop a high-throughput
425 screening method for the quantification of *Salmonella* in cell culture infections, we used the
426 VideoScan Platform developed used in our lab for nucleic acid detection or quantification, *E.*
427 *coli* quantification in cell culture adhesion assays, quantification of bacterial adhesion to
428 proteins, and automated detection of autoantibodies in cell-based assays (23, 43–45). The new
429 VideoScan module leaverges the detection of and screening of adherent and invasive GFP-
430 expressing bacteria to provide an output of bacterial numbers adhering to and invading a cell
431 line. This assay has advantages over other bacterial quantification methods in that it is
432 simpler, avoids complicated staining procedures involving membrane permeabilization,
433 avoids the often laborious selection of suitable antibodies for different serovars and problems
434 arising from non-specific antibody binding, and there is no need to perform dilution series and
435 CFU determinations on agar plates (46). The only disadvantage of this approach is the
436 requirement for bacterial transformation with a plasmid encoding the GFP protein, where it
437 could be difficult to find an antibiotic selection marker for antibiotic-resistant isolates (47).
438 This approach was used to test if 127 *Salmonella* isolates, representing serovars with
439 different host ranges, were better at infecting cells originating from the associated host.
440 However, using this model this was shown not to be the case. This might be because there are
441 many other host factors related to *Salmonella* host specificity, that were not present in our
442 simple model of epithelial cell infection (48).
443 In an effort to explain the differences in infectivity among isolates, we focused our efforts to
444 analyse and identify the possible underlying genetic explanations for the outcomes in ourof
445 the cell line infection assays in isolates that showed altered infection phenotypes in
446 comparison to the majority of isolates screened in each group. Though *Salmonella* virulence
447 plasmid (pSV) was previously associated with septicaemia and enteritis (49), we were not

able to find any association between altered infection phenotype and the presence of pSV. However, we were able to find naturally occurring mutations present in the genomes of clinical isolates of *Salmonella*. Interestingly, a majority of the mutations we found were not associated with any factors directly responsible for invasion or adhesion. We did not identify any mutations in genes encoding T3SS, Rck or PagN that would explain the altered infection phenotypes, nor in genes encoding adhesins, except for a single base substitution in the *fimH* gene leading to premature translation termination in one of the isolates with a high-infection phenotype. Usually, loss of type 1 fimbriae (T1F) expression (*fimH* gene codes for tip adhesin of T1F), which this deletion confers, generally results in lower cell adhesion (50). Noteworthy is the fact that bacteria for adhesion assays with T1F are usually grown under different conditions from those used in our screens (12, 51).

When we looked at the function of genes with natural mutations, we noticed that they are related to cell membrane/surface structures and cell wall or surface structure expression regulation. There are reports which link pseudogene formation in cell membrane/surface structures with host adaptation (52, 53). These could explain the presence of natural mutations in host-restricted (*S.* Gallinarum) and host-adapted serovars (*S.* Dublin and *S.* Choleraesuis). Other explanations for natural mutations could be that these isolates were isolated from asymptomatic carriers and the presence of natural mutations might increase long-term survival within the host without recognition by the immune system. Similar findings were ascertained by Klemm et al., (2016), where the formation of pseudogenes in membrane proteins have been reported in *S.* Enteritidis isolates of one strain sampled multiple times over 15 years from one patient (54). Unfortunately, we had no information about the host's clinical status at the time of sampling, and therefore we cannot confirm or rule out this possibility.

Another possibility that could explain the occurrence of altered infection phenotypes and natural mutations is the mutagenic properties of antibiotics or other xenobiotics. Multiple

studies have shown that various classes of antibiotics can induce stress and mutations at subinhibitory concentrations (55–57). These mutations are the effect of adaptive bacterial response to antimicrobials or a consequence of direct interaction of antibiotics with bacterial DNA (58). When we checked for the presence of antibiotic resistance genes, nine out of 30 *Salmonella* isolates had at least one antibiotic resistance gene, confirming a possible exposure to antibiotics. Antibiotic-induced mutations have a deleterious effect, such that the bacteria bearing them would have decreased fitness and more negative selection in comparison to wildtype isolates. This negative selection could perhaps explain why the natural mutations that we observe in our strain collection can be rarely found in GenBank. To confirm this hypothesis in future, an antibiotic-driven mutant library could be created and tested for infection of epithelial cells.

Contribution of mutations to altered infection phenotype was functionally validated for two genes: *dksA* and *sanA*. Gene *dksA* was selected because its contribution to host cell adhesion and invasion was known. Little was known about the influence of *sanA* on *Salmonella* host cell adhesion, therefore they were chosen for further investigation. Complementation of the mutation found in *dksA* led to an increase in growth rate in minimal medium and host cell infectivity. Though the importance of the amino acid at position 88 in DksA for growth in minimal media has been described previously (59), a link to cell line invasion has not been described so far. Deletion of *sanA* was previously associated with increased susceptibility to vancomycin in *E. coli* (60). One of our isolates with altered infection phenotypes had a 10 nucleotide deletion in *sanA*. We therefore exchanged the mutation with the wild-type sequence and determined whether this would lead to rescue of certain functions. Indeed, complementation of the mutation led to a decrease in resistance to vancomycin, and further influenced adhesion and invasion. Though the mode of action for *dksA* in adhesion and invasion was described for *S.* Typhimurium in previous studies (59, 61,

62), there is no information about the role or possible contribution of *dksA* sequence variation and *sanA* to bacterial adhesion or invasion.

We decided to get a global overview of expression profiles for the isolates with altered infection phenotypes, so we conducted a transcriptomic analysis, which highlighted the importance of T3SS expression levels in the determination of an isolate's infection phenotype. Similar differences were observed in the study of Shah (2014), where 38 genes from SPI-1 showed reduced expression in less virulent *S.* Enteritidis isolates in comparison to more virulent *S.* Enteritidis (63). Three isolates with low infection phenotype included in this analysis had mutations in *aroA*, *ompD*, *yidR* and *sirA*. Though the contribution of *aroA* and *sirA* in the regulation of genes associated with T3SS has been shown previously and there is information about a possible connection between T3SS and *ompD* expression levels (64–66), there is so far no information about the *yidR* gene. One strain with low infection phenotype and a mutation in the *sanA* gene had a gene expression profile similar to highly infecting phenotype. Since knowledge about this gene is scant, an explanation for this observation requires further study.

Our analysis of genomes and transcriptomes of the isolates with altered infection phenotypes allowed us to find explanations for lowered infection of epithelial cells. However, we were not able to identify the factors contributing to highly infecting phenotypes of the selected *Salmonella* isolates. It is possible that other factors, like methylation, influence higher virulence of these isolates, as shown before in the case of *Salmonella* (67). A detailed investigation of these additional factors is out of the scope of present investigation.

To summarize, this work was an exploration of what we can learn from the observing the interaction of clinical isolates of *Salmonella* with host cells of intestinal epithelial origin. We have shown that the majority of isolates have similar infection phenotypes and isolates with altered infection phenotypes provide a source for determination of new genes or new gene

variants influencing epithelial infection, including novel factors associated with adhesion and invasion of epithelial cells by *Salmonella*. We identified two genes, *sanA* and *yidR*, that were not previously associated with adhesion or invasion of host cells. We consider these genes to be interesting candidates for further investigations within the context of *Salmonella* virulence. Our study reveals new questions regarding the origin of such mutations and their contribution to pathogen evolution and pathogenicity, and suggest that the approach used here with larger strain collections belonging to the same serovar, clinical data and the use of genome-wide association studies could provide important insights into *Salmonella* host cell interactions.

5. Acknowledgements

This work was supported by the Federal Ministry of Education and Research, Germany (BMBF InnoProfile-Transfer 03IPT611X) and by the National Science Centre, Poland (Miniatura 2, DEC-2018/02/X/NZ6/00628). Derek Pickard was generously funded by the Wellcome Trust.

6. References

1. Dougan G, Baker S. 2014. *Salmonella enterica* serovar Typhi and the pathogenesis of typhoid fever. Annu Rev Microbiol 68:317–336.

2. Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, Praet N, Bellinger DC, de Silva NR, Gargouri N, Speybroeck N, Cawthorne A, Mathers C, Stein C, Angulo FJ, Devleesschauwer B, World Health Organization Foodborne Disease Burden Epidemiology Reference Group. 2015. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLoS Med 12:e1001923.

3. Barrow PA, Methner U. 2013. *Salmonella* in Domestic Animals. CABI.

4. Uzzau S, Brown DJ, Wallis T, Rubino S, Leori G, Bernard S, Casadesús J, Platt DJ, Olsen JE. 2000. Host adapted serotypes of *Salmonella enterica*. Epidemiol Infect 125:229–255.

5. Stevens MP, Humphrey TJ, Maskell DJ. 2009. Molecular insights into farm animal and zoonotic *Salmonella* infections. Philos Trans R Soc Lond B Biol Sci 364:2709–2723.

6. Grzymajlo K, Kuzminska-Bajor M, Jaworski J, Dobryszycki P, Ugorski M. 2010. The high-adhesive properties of the FimH adhesin of *Salmonella enterica* serovar Enteritidis are determined by a single F118S substitution. Microbiol Read Engl 156:1738–1748.

7. Chiu C-H, Su L-H, Chu C. 2004. *Salmonella enterica* Serotype Choleraesuis: Epidemiology, Pathogenesis, Clinical Disease, and Treatment. Clin Microbiol Rev 17:311–322.

8. Nielsen LR. 2013. Review of pathogenesis and diagnostic methods of immediate relevance for epidemiology and control of *Salmonella* Dublin in cattle. Vet Microbiol 162:1–9.

9. Shivaprasad HL. 2000. Fowl typhoid and pullorum disease. Rev Sci Tech Int Off Epizoot 19:405–424.

10. Sansonetti PJ. 2004. War and peace at mucosal surfaces. Nat Rev Immunol 4:953–964.

11. Velge P, Wiedemann A, Rosselin M, Abed N, Boumart Z, Chaussé AM, Grépinet O, Namdari F, Roche SM, Rossignol A, Virlogeux-Payant I. 2012. Multiplicity of *Salmonella* entry mechanisms, a new paradigm for Salmonella pathogenesis. MicrobiologyOpen 1:243–258.

12. Kolenda R, Ugorski M, Grzymajlo K. 2019. Everything You Always Wanted to Know About *Salmonella* Type 1 Fimbriae, but Were Afraid to Ask. Front Microbiol 10:1017.

13. Wagner C, Hensel M. 2011. Adhesive mechanisms of *Salmonella enterica*. Adv Exp Med Biol 715:17–34.

14. Lostroh CP, Lee CA. 2001. The *Salmonella* pathogenicity island-1 type III secretion system. Microbes Infect 3:1281–1291.

15. Lambert MA, Smith SGJ. 2008. The PagN protein of *Salmonella enterica* serovar Typhimurium is an adhesin and invasin. BMC Microbiol 8:142.

16. Rosselin M, Virlogeux-Payant I, Roy C, Bottreau E, Sizaret P-Y, Mijouin L, Germon P, Caron E, Velge P, Wiedemann A. 2010. Rck of *Salmonella enterica*, subspecies enterica serovar enteritidis, mediates zipper-like internalization. Cell Res 20:647–664.

17. Coburn B, Sekirov I, Finlay BB. 2007. Type III secretion systems and disease. Clin Microbiol Rev 20:535–549.

18. Ghatak S, King ZA, Sastry A, Palsson BO. 2019. The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. Nucleic Acids Res 47:2446–2454.

19. Frömmel U, Böhm A, Nitschke J, Weinreich J, Groß J, Rödiger S, Wex T, Ansorge H, Zinke O, Schröder C, Roggenbuck D, Schierack P. 2013. Adhesion patterns of commensal and pathogenic *Escherichia coli* from humans and wild animals on human and porcine epithelial cell lines. Gut Pathog 5:31.

20. Schierack P, Rödiger S, Kuhl C, Hiemann R, Roggenbuck D, Li G, Weinreich J, Berger E, Nolan LK, Nicholson B, Römer A, Frömmel U, Wieler LH, Schröder C. 2013.

592 Porcine *E. coli* : Virulence-Associated Genes, Resistance Genes and Adhesion and
593 Probiotic Activity Tested by a New Screening Method. PLOS ONE 8:e59242.

594 21. Valdivia RH, Falkow S. 1996. Bacterial genetics by flow cytometry: rapid isolation of
595 Salmonella typhimurium acid-inducible promoters by differential fluorescence
596 induction. Mol Microbiol 22:367–378.

597 22. Schierack P, Nordhoff M, Pollmann M, Weyrauch KD, Amasheh S, Lodemann U, Jores
598 J, Tachu B, Kleta S, Blikslager A, Tedin K, Wieler LH. 2006. Characterization of a
599 porcine intestinal epithelial cell line for in vitro studies of microbial pathogenesis in
600 swine. Histochem Cell Biol 125:293–305.

601 23. Rödiger S, Schierack P, Böhm A, Nitschke J, Berger I, Frömmel U, Schmidt C, Ruhland
602 M, Schimke I, Roggenbuck D, Lehmann W, Schröder C. 2013. A highly versatile
603 microscope imaging technology platform for the multiplex real-time detection of
604 biomolecules and autoimmune antibodies. Adv Biochem Eng Biotechnol 133:35–74.

605 24. Wingett SW, Andrews S. 2018. FastQ Screen: A tool for multi-genome mapping and
606 quality control. F1000Research 7:1338.

607 25. Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, Otto TD, Keane JA.
608 2016. Robust high-throughput prokaryote de novo assembly and improvement pipeline
609 for Illumina data. Microb Genomics 2:e000083.

610 26. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl
611 30:2068–2069.

27. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinforma Oxf Engl 31:3691–3693.

28. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5.

29. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259.

30. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 17:238.

31. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361.

32. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895–3903.

33. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

34. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28:464–469.

633 35. Khetrapal V, Mehershahi K, Rafee S, Chen S, Lim CL, Chen SL. 2015. A set of
634 powerful negative selection systems for unmodified Enterobacteriaceae. Nucleic Acids
635 Res 43:e83.

636 36. Sprouffske K, Wagner A. 2016. Growthcurver: an R package for obtaining interpretable
637 metrics from microbial growth curves. BMC Bioinformatics 17:172.

638 37. Stead MB, Agrawal A, Bowden KE, Nasir R, Mohanty BK, Meagher RB, Kushner SR.
639 2012. RNAsnap™: a rapid, quantitative and inexpensive, method for isolating total RNA
640 from bacteria. Nucleic Acids Res 40:e156.

641 38. Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-
642 time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif
643 25:402–408.

644 39. Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and
645 better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res
646 47:e47.

647 40. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for
648 differential expression analysis of digital gene expression data. Bioinforma Oxf Engl
649 26:139–140.

650 41. R Development Core Team. R Development Core Team (2013). R: A language and
651 environment for statistical computing. R Foundation for Statistical Computing, Vienna,
652 Austria. ISBN 3-900051-07-0, URL http://www.R-project.org. R Foundation for
653 Statistical Computing, Vienna, Austria.

654 42. Wickham H. 2010. A Layered Grammar of Graphics. J Comput Graph Stat 19:3–28.

43. Ali A, Kolenda R, Khan MM, Weinreich J, Li G, Wieler LH, Tedin K, Roggenbuck D, Schierack P. 2020. Identification of Avian Pathogenic *Escherichia coli*'s novel genes responsible for adhesion to chicken and human cell lines. Appl Environ Microbiol https://doi.org/10.1128/AEM.01068-20.

44. Kolenda R, Burdukiewicz M, Schiebel J, Rödiger S, Sauer L, Szabo I, Orłowska A, Weinreich J, Nitschke J, Böhm A, Gerber U, Roggenbuck D, Schierack P. 2018. Adhesion of *Salmonella* to Pancreatic Secretory Granule Membrane Major Glycoprotein GP2 of Human and Porcine Origin Depends on FimH Sequence Variation. Front Microbiol 9:1905.

45. Sowa M, Kolenda R, Baumgart DC, Pratschke J, Papp M, Tornai T, Suchanski J, Bogdanos DP, Mytilinaiou MG, Hammermann J, Laass MW, Conrad K, Schramm C, Franke A, Roggenbuck D, Schierack P. 2018. Mucosal Autoimmunity to Cell-Bound GP2 Isoforms Is a Sensitive Marker in PSC and Associated With the Clinical Phenotype. Front Immunol 9:1959.

46. Wu J, Pugh R, Laughlin RC, Andrews-Polymenis H, McClelland M, Bäumler AJ, Adams LG. 2014. High-throughput assay to phenotype *Salmonella enterica* Typhimurium association, invasion, and replication in macrophages. J Vis Exp JoVE e51759.

47. Van Puyvelde S, Pickard D, Vandelannoote K, Heinz E, Barbé B, de Block T, Clare S, Coomber EL, Harcourt K, Sridhar S, Lees EA, Wheeler NE, Klemm EJ, Kuijpers L, Mbuyi Kalonji L, Phoba M-F, Falay D, Ngbonda D, Lunguya O, Jacobs J, Dougan G, Deborggraeve S. 2019. An African *Salmonella* Typhimurium ST313 sublineage with extensive drug-resistance and signatures of host adaptation. Nat Commun 10:4280.

678 48. Bäumler A, Fang FC. 2013. Host Specificity of Bacterial Pathogens. Cold Spring Harb
679 Perspect Med 3.

680 49. Silva C, Puente JL, Calva E. 2017. *Salmonella* virulence plasmid: pathogenesis and
681 ecology. Pathog Dis https://doi.org/10.1093/femspd/ftx070.

682 50. Grzymajlo K, Ugorski M, Suchanski J, Kedzierska AE, Kolenda R, Jarzab A,
683 Biernatowska A, Schierack P. 2017. The Novel Type 1 Fimbriae FimH Receptor
684 Calreticulin Plays a Role in *Salmonella* Host Specificity. Front Cell Infect Microbiol
685 7:326.

686 51. Klasa B, Kędzierska AE, Grzymajło K. 2020. Pre-Growth Culture Conditions Affect
687 Type 1 Fimbriae-Dependent Adhesion of *Salmonella*. Int J Mol Sci 21.

688 52. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, Seth-Smith
689 HMB, Barquist L, Stedman A, Humphrey T, Wigley P, Peters SE, Maskell DJ, Corander
690 J, Chabalgoity JA, Barrow P, Parkhill J, Dougan G, Thomson NR. 2015. Patterns of
691 genome evolution that have accompanied host adaptation in *Salmonella*. Proc Natl Acad
692 Sci U S A 112:863–868.

693 53. Okoro CK, Barquist L, Connor TR, Harris SR, Clare S, Stevens MP, Arends MJ, Hale C,
694 Kane L, Pickard DJ, Hill J, Harcourt K, Parkhill J, Dougan G, Kingsley RA. 2015.
695 Signatures of Adaptation in Human Invasive *Salmonella* Typhimurium ST313
696 Populations from Sub-Saharan Africa. PLOS Negl Trop Dis 9:e0003611.

697 54. Klemm EJ, Gkrania-Klotsas E, Hadfield J, Forbester JL, Harris SR, Hale C, Heath JN,
698 Wileman T, Clare S, Kane L, Goulding D, Otto TD, Kay S, Doffinger R, Cooke FJ,
699 Carmichael A, Lever AM, Parkhill J, MacLennan CA, Kumararatne D, Dougan G,

700 Kingsley RA. 2016. Emergence of host-adapted *Salmonella* Enteritidis through rapid
701 evolution in an immunocompromised host. Nat Microbiol 1.

702 55. Kang TM, Yuan J, Nguyen A, Becket E, Yang H, Miller JH. 2012. The aminoglycoside
703 antibiotic kanamycin damages DNA bases in *Escherichia coli*: caffeine potentiates the
704 DNA-damaging effects of kanamycin while suppressing cell killing by ciprofloxacin in
705 *Escherichia coli* and *Bacillus anthracis*. Antimicrob Agents Chemother 56:3216–3223.

706 56. Gutierrez A, Laureti L, Crussard S, Abida H, Rodríguez-Rojas A, Blázquez J, Baharoglu
707 Z, Mazel D, Darfeuille F, Vogel J, Matic I. 2013. β-Lactam antibiotics promote bacterial
708 mutagenesis via an RpoS-mediated reduction in replication fidelity. Nat Commun
709 4:1610.

710 57. Song LY, Goff M, Davidian C, Mao Z, London M, Lam K, Yung M, Miller JH. 2016.
711 Mutational Consequences of Ciprofloxacin in *Escherichia coli*. Antimicrob Agents
712 Chemother 60:6165–6172.

713 58. Perigolo de Oliveira M, Constant J-F, Peuchmaur M, Pitta I, Décout J-L. 2013.
714 Antibiotic drugs aminoglycosides cleave DNA at abasic sites: shedding new light on
715 their toxicity? Chem Res Toxicol 26:1710–1719.

716 59. Azriel S, Goren A, Rahav G, Gal-Mor O. 2016. The Stringent Response Regulator DksA
717 Is Required for *Salmonella enterica* Serovar Typhimurium Growth in Minimal Medium,
718 Motility, Biofilm Formation, and Intestinal Colonization. Infect Immun 84:375–384.

719 60. Rida S, Caillet J, Alix JH. 1996. Amplification of a novel gene, *sanA*, abolishes a
720 vancomycin-sensitive defect in *Escherichia coli*. J Bacteriol 178:94–102.

61. Webb C, Moreno M, Wilmes-Riesenberg M, Curtiss R, Foster JW. 1999. Effects of DksA and ClpP protease on sigma S production and virulence in *Salmonella* typhimurium. Mol Microbiol 34:112–123.

62. Rice CJ, Ramachandran VK, Shearer N, Thompson A. 2015. Transcriptional and Post-Transcriptional Modulation of SPI1 and SPI2 Expression by ppGpp, RpoS and DksA in *Salmonella enterica* sv Typhimurium. PloS One 10:e0127523.

63. Shah DH. 2014. RNA sequencing reveals differences between the global transcriptomes of Salmonella enterica serovar enteritidis strains with high and low pathogenicities. Appl Environ Microbiol 80:896–906.

64. Ahmer BM, van Reeuwijk J, Watson PR, Wallis TS, Heffron F. 1999. *Salmonella* SirA is a global regulator of genes mediating enteropathogenesis. Mol Microbiol 31:971–982.

65. Pfeiffer V, Sittka A, Tomer R, Tedin K, Brinkmann V, Vogel J. 2007. A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella* core genome. Mol Microbiol 66:1174–1191.

66. Felgner S, Frahm M, Kocijancic D, Rohde M, Eckweiler D, Bielecka A, Bueno E, Cava F, Abraham W-R, Curtiss R, Häussler S, Erhardt M, Weiss S. 2016. *aroA*-Deficient *Salmonella enterica* Serovar Typhimurium Is More Than a Metabolically Attenuated Mutant. mBio 7.

67. Heithoff DM, Sinsheimer RL, Low DA, Mahan MJ. 1999. An essential role for DNA adenine methylation in bacterial virulence. Science 284:967–970.

68. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645.

743 69. Fardini Y, Chettab K, Grépinet O, Rochereau S, Trotereau J, Harvey P, Amy M,
744 Bottreau E, Bumstead N, Barrow PA, Virlogeux-Payant I. 2007. The YfgL Lipoprotein
745 Is Essential for Type III Secretion System Expression and Virulence of *Salmonella*
746 *enterica* Serovar Enteritidis. Infect Immun 75:358–370.

747

748

749

750

751

752 Table 1. List of bacteria used in this study

Strains	Description (Group name)	Reference numbers	Source
Escherichia coli XL1Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIq ZΔM15 Tn10 (Tetr)]*		Agilent Technologies
Salmonella enterica subsp. *enterica* serovar Typhimurium	20 strains, isolated from human diarrhea cases (Typhimurium-human)	5717-5725, 6181- 6191	Mydlak/Thorausch Diagnostic Laboratory, Cottbus
Salmonella enterica subsp. *enterica* serovar Typhimurium	20 strains, isolated from porcine faeces samples (Typhimurium-pig)	5726-5745	BfR, Berlin
Salmonella enterica subsp. *enterica* serovar Enteritidis	20 strains, isolated from human diarrhea cases (Enteritidis-human)	5702-5710, 6170-6180	Mydlak/Thorasch Diagnostic Laboratory, Cottbus
Salmonella enterica subsp. *enterica* serovar Enteritidis	14 strains, isolated from chicken faeces samples (Enteritidis-chicken)	6192-6195,6197, 6198, 6200-6207	German Federal Institute for Risk assessment (BfR), Berlin
Salmonella enterica subsp. *enterica* serovar Gallinarum	19 strains, isolated from chicken faeces samples (Gallinarum)	5766-5771, 5773-5785	BfR, Berlin
Salmonella enterica subsp. *enterica* serovar Dublin	20 strains, isolated from bovine faeces samples (Dublin)	5746-5765	BfR, Berlin
Salmonella enterica	15 strains, isolated from	6053, 6054,	BfR, Berlin

subsp. *enterica* serovar Choleraesuis	various sources-porcine faeces, wild boars, meat products, reptile (Choleraesuis)	6138-6150	
Salmonella enterica subsp. *enterica* serovar Typhimurium SL1344			Free University, Berlin

753
754
755
756
757
758

759 Table 2.

760 List of primers used in this study.

Primer name	Primer sequence (5'->3')	Gene	Source
pFVP25.1KmRev	CTGTCAGACCAAGTTTACTCATATATAC	*kanR*	This study
KanpFVP25Fwd	TGAGTAAACTTGGTCTGACAGTCAGAAGAACTCGTCAAGAAG	*kanR*	This study
KanpFVP25Rev	TAATATTGAAAAAGGAAGAGTATGATTGAACAAGATGGATTG	*kanR*	This study
pFVP25.1KmFwd	ACTCTTCCTTTTTCAATATTATTG	*kanR*	This study
dksAdelNEGfor	ATGCAAGAAGGGCAAAACCGTAAAACATCGTCCCTGAGTATTCTCGCCAT GCTTGCAGTGGGCTTACATG	*dksA*	This study
dksAdelNEGrev	TTAACCCGCCATCTGTTTTTCGCGAATTTCAGCCAGCGTTTTGCAGTCGA TCAGAGCAGGATCGACGTCC	*dksA*	This study
dksAfor	ATGCAAGAAGGGCAAAACCG	*dksA*	This study
dksArev	TTAACCCGCCATCTGTTTTTCG	*dksA*	This study
dksA100UpstreamFor	ACAGGGTTGTCAAGTGTTACG	*dksA*	This study
dksA100DownstreamRev	TAACGAGCCGAAATGCAGTTC	*dksA*	This study
dksAinternalRev	AGTGCGATCGACTTCATCC	*dksA*	This study
k2	CGGTGCCCTGAATGAACTGC	*kanR*	(68)
sanAdelNEGfor	ATGTTAAAGCGCGTGTTTTACAGCCTGTTGGTCCTGGTAGGCTTGCTGCT GCTTGCAGTGGGCTTACATG	*sanA*	This study
sanAdelNEGrev	TCATTTCCCTTTTTTCTTTTCCAGTTCAAGCAATTGTTCCGGCGTAACTG TCAGAGCAGGATCGACGTCC	*sanA*	This study
sanAfor	ATGTTAAAGCGCGTGTTTTAC	*sanA*	This stud

			y
sanArev	TCATTTCCCTTTTTTCTTTTCCAG	*sanA*	This study
sanA100UpstreamFor	CGATACAAGGGAAATCATGCTG	*sanA*	This study
sanA100DownstreamRev	TTCCAGGCCTCACGGAAG	*sanA*	This study
sanAinternalRev	GCCCTGGATACGATAACGA	*sanA*	This study
tufAqPCRfor	TGTTCCGCAAACTGCTGGACG	*tufA*	(69)
tufAqPCRrev	ATGGTGCCCGGCTTAGCCAGTA	*tufA*	(69)
dksAqPCRfor1	TGAAGCATGGCGTAATCAACTC	*dksA*	This study
dksAqPCRrev1	TCCAGGCTAAACTCCTCTTCC	*dksA*	This study
sanAqPCRfor1	AAACAGCGCCCTATATCTATGAC	*sanA*	This study
sanAqPCRrev1	TGATTAATGACACCCTTGCGA	*sanA*	This study

761

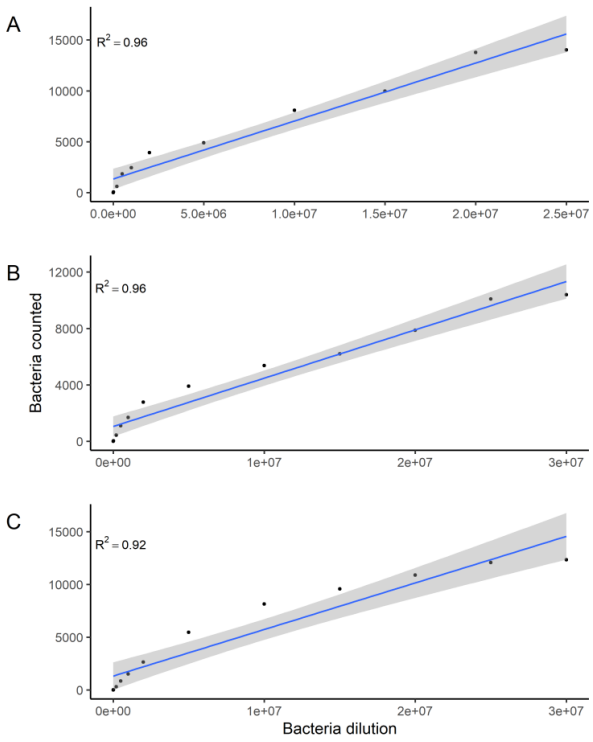
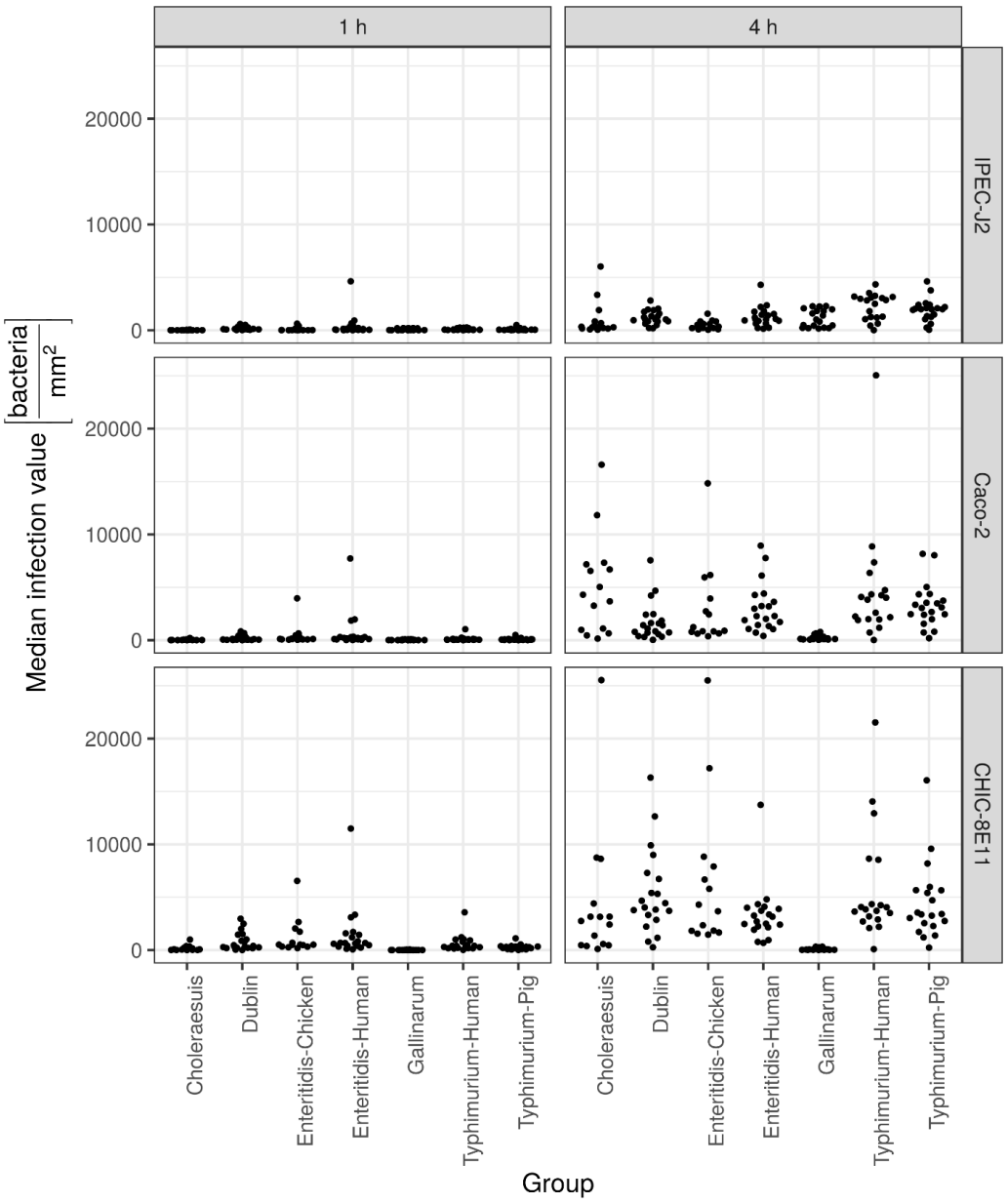


Figure 1. Dilution series of *Salmonella* in cell line infection assays.

This assay was performed to determine the linear range of GFP-expressing *S.* Typhimurium SL1344 in infection assays with IPEC-J2 (A), Caco-2 (B) and CHIC-8E11 (C) cell lines in a 96-well plate format for 1 h. Images were automatically taken by the VideoScan instrument. Black dots represent a median value from three assays in triplicates for each bacterial dilution. Blue lines connecting dots are smoothed trend lines and grey areas around the lines are 0.95 confidence interval of the linear regression.



771

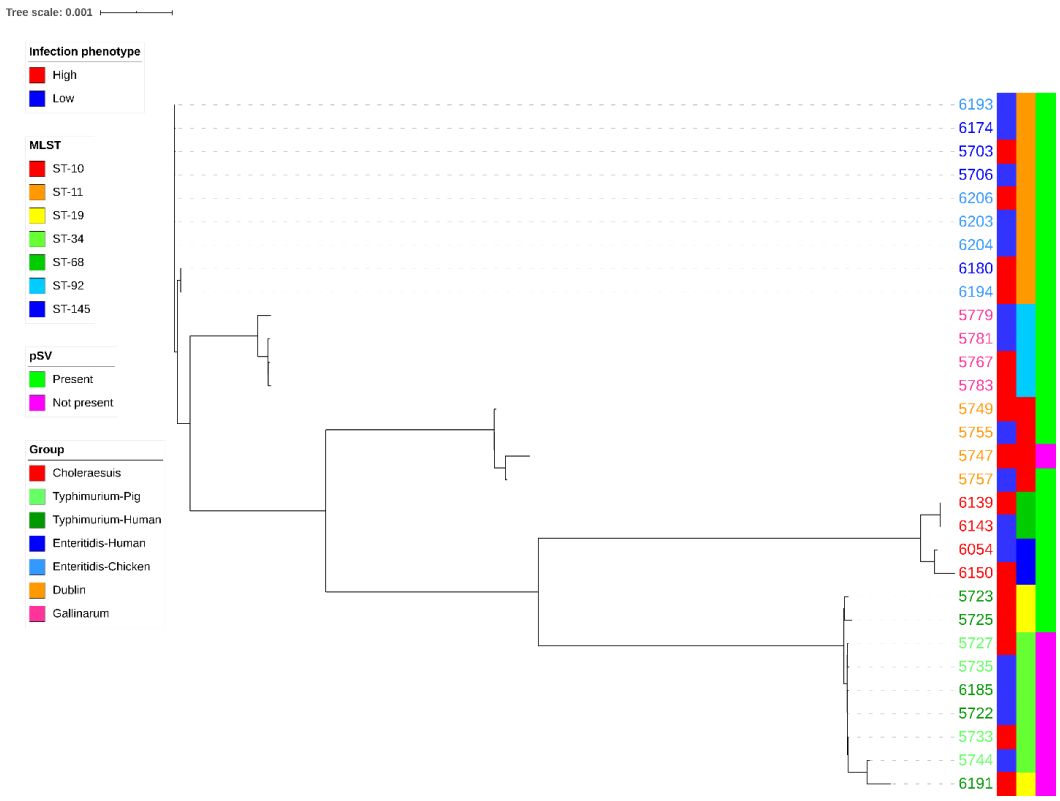
772 **Figure 2. *Salmonella* infection of IPEC-J2, Caco-2 and CHIC-8E11 cell lines.**

773 127 *Salmonella* isolates from *S.* Typhimurium of human (Typhimurium-Human) and pig

774 (Typhimurium-Pig) origin, *S.* Enteritidis of human (Enteritidis-Human) and chicken (Enteritidis-

775 Chicken) origin, *S.* Choleraesuis (Choleraesuis), *S.* Dublin (Dublin), *S.* Gallinarum (Gallinarum)
776 were compared for IPEC-J2, Caco-2 and CHIC-8E11 cell line infection after 1 h and 4 h of
777 incubation. Each dot represents a median infection value from three measurements in triplicates
778 for one strain. Results are shown as number of bacteria per mm^2.

779



780

781 **Figure 3. Core genome tree of *Salmonella* with different infection phenotypes**

782 Annotated genomes were analysed with Roary. Next, the core genome alignment of 3473 genes
783 from 30 genomes was used to generate a tree with FastTree 2.1. Information about infection
784 phenotype, group, MLST sequence types and presence of pSV plasmid were added with the use
785 of iTOL. Groups and isolate numbers used in analysis: Choleraesuis- 6054, 6139, 6143, 6150;
786 Typhimurium-Pig- 5727, 5733, 5735, 5744; Typhimurium-Human- 5722, 5723, 5725, 6185,

787 6191; Enteritidis-Human- 5703, 5706, 6174, 6180; Enteritidis-Chicken- 6193, 6194, 6203, 6204,
788 6206, Dublin- 5747, 5749, 5755, 5757; Gallinarum- 5767, 5779, 5781, 5783.
789
790 Table 3. Candidate genes responsible for altered infection phenotype with a description of their
791 function (where possible), mutation and presence of genomes with the same changes in the
792 GenBank database (GenBank query was made with use of RefSeq Genomes Database).
793 Abbreviations: NN- nucleotide, AA- amino acid, Tym- *S.* Typhimurium, cds- coding sequence

Serovar	Gene	Function	Mutation	GenBank query*
S. Enteritidis	*dksA*	Required for growth in minimal medium and intestinal colonization; mediates adaptation to oxidative and nitrosative stress.	Two AA deletion at position 87 and 88 in isolate 5706	Yes (54 sequences, 14 sequences with mutations in the same region)
S. Enteritidis	*ompD/nmpC*	Porin OmpD, mediates binding of Tym to T84 and U937 cells	Deletion of starting 300 bp of cds in isolate 6174	No
S. Enteritidis	*yidR*	Putative ATP/GTP-binding protein; mutant has reduced binding to/persistence in lettuce.	Stop codon at position 317 in isolate 6174	Yes (4 genomes, 21.10.2019)
S. Enteritidis	*sanA*	Vancomycin high-temperature exclusion protein; required for bile tolerance in *S*. Typhi.	Deletion of 10 NN in isolate 6203	Yes (1 genome, 14 NN deletions in the same region, 21.10.2019)

S. Enteritidis	*rfaL (waaL)*	O-antigen ligase; in Tym affects motility.	12 NN deletion in isolate 6193	No (21.10.2019)
S. Typhimurium	*sirA*	Invasion response-regulator; global regulator of genes mediating enteropathogenesis.	14 NN deletion in isolate 5735	Yes (2 genomes, 21.10.2019)
S. Typhimurium	*aroA*	3-phosphoshikimate 1-carboxyvinyltransferase; mutation leads to auxotrophy for aromatic amino acids.	78 NN deletion in isolate 6185	No (21.10.2019)
S. Typhimurium	*fimH*	Adhesin of type 1 fimbriae.	Stop codon at position 80 in isolate 6191	Yes (13 genomes, 21.10.2019)
S. Gallinarum	*lrhA*	Transcriptional regulator; negative regulator of flagella expression in early cell growth phases in Tym.	3 NN deletion (at 100 position), 1 AA (Ala), in isolate 5781	Yes (7 genomes, 21.10.2019)
S. Gallinarum	*rcsD*	Phosphotransfer intermediate protein in a two-component regulatory system with RcsBC; is implicated in the control of capsule and flagella synthesis and biofilm formation.	5 NN deletion in isolate 5781	No
S. Dublin	*ompW*	Outer membrane protein W; mediates methyl viologen	1 NN deletion in isolate 5755, 5757	No

		(paraquat) efflux in Tym.		
S. Dublin	*mpl*	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase; peptidoglycan synthesis.	Deletion after codon 423 in isolate 5757	No
S. Choleraesuis	*rtsA*	Increases expression of the invasion genes by inducing *hilA, hilD, hilC* genes and the *invF* operon.	1 SNP that leads to change from Arg to Gln in isolate 6143.	No
S. Choleraesuis	*wza*	Polysaccharide (colanic acid) export protein Wza; involved in exopolysaccharide (EPS) production.	1 SNP in codon 171 that leads to change for stop codon in isolate 6054	No

794

795

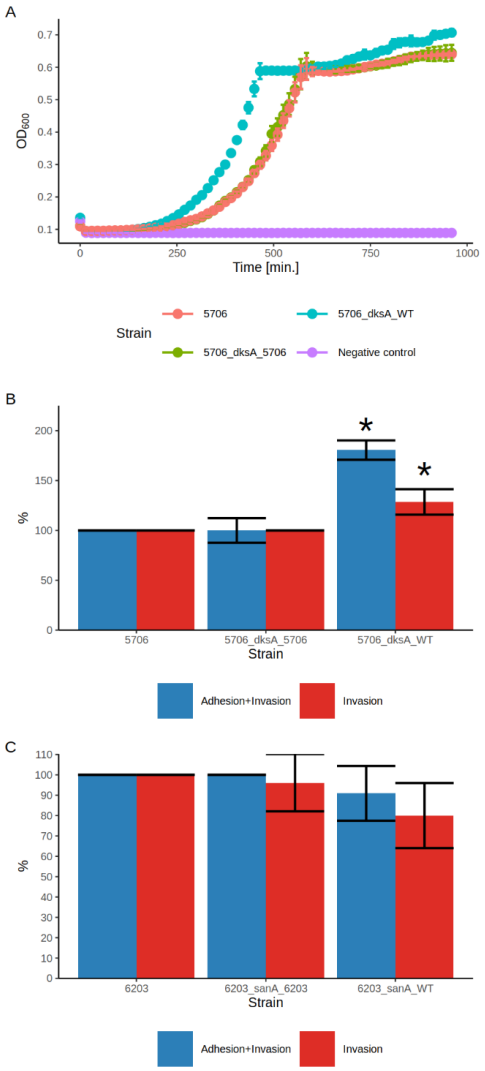


Figure 4. Phenotypic assays for *dksA* ans *sanA* variants.

A) Growth curve of *S.* Enteritidis 5706 isogenic strains in M9 minimal medium supplemented with 0.2% glucose. Dots represent a median value from three assays in triplicates for each time point. B) Gentamycin protection assay with the use of *S.* Enteritidis 5706 isogenic strains and Caco-2 cells. Strains used in A) and B): 5706- isolate with mutated *dksA* variant,

5706_dksA_5706- isolate with mutated *dksA* variant introduced by negative selection system, 5706_dksA_WT- isolate with wildtype *dksA* variant introduced by negative selection system.

C) Gentamycin protection assay with the use of *S.* Enteritidis 6203 isogenic strains and Caco-2 cells. Strains used in C) 6203- isolate with mutated *sanA* variant, 6203_sanA_6203- isolate with mutated *sanA* variant introduced by negative selection, 6203_sanA_WT- isolate with wildtype *sanA* variant introduced by negative selection. Error bars represent the median absolute deviation. Statistically significant results in B) were marked with "*" (p=0.05).

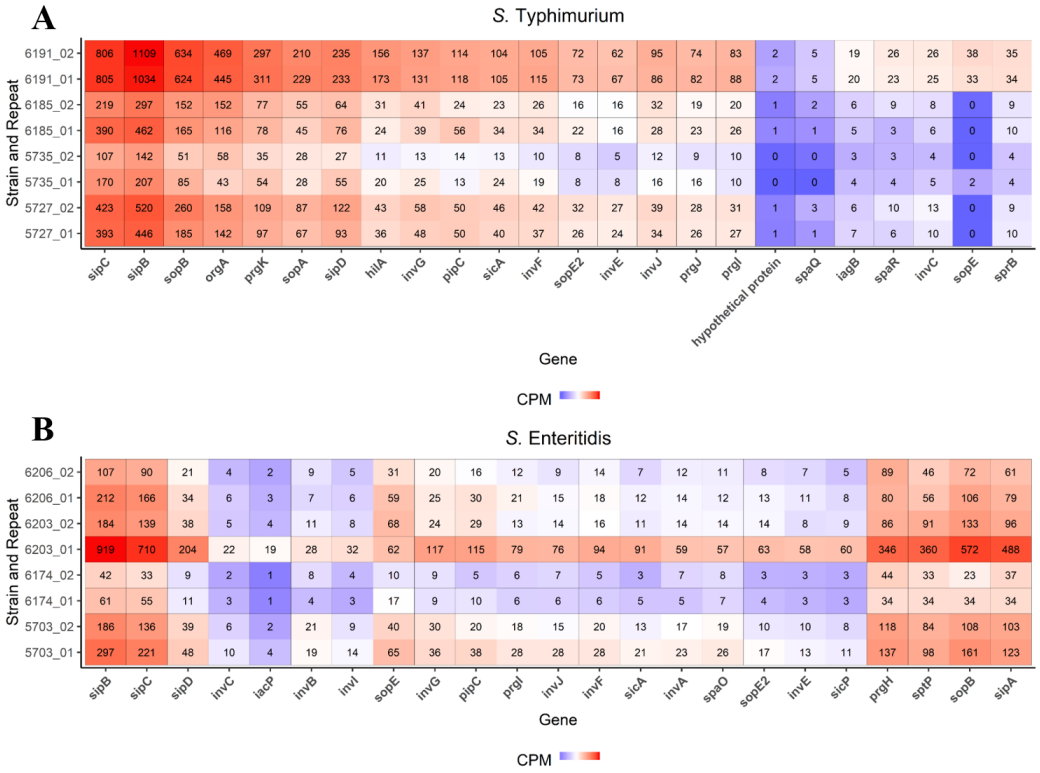


Figure 5. Differentially expressed genes associated with epithelial cell invasion.

812 Heatmap with expression levels of genes associated with invasion of epithelial cells in four *S.*
813 Typhimurium (A) and four *S.* Enteritidis isolates (B). Strain number and repeat are shown on the
814 y-axis. Isolates 5703, 5727, 6191 and 6206 represent high-infecting cell line phenotypes while
815 isolates 5735, 6174, 6185 and 6203 represent low-infecting phenotypes. Names of genes DE
816 expressed are shown on the x-axis. Color gradient is proportional to counts per million (CPM) for
817 each gene in each repeat. Number of CPM for each gene and repeat are written in rectangles.